

Incêndios Australianos e o Papel das Condições Climáticas em Seus Números

Gabriel B. Gutierrez

Faculdade de Computação – Universidade Federal do Mato Grosso do Sul (UFMS)
Campo Grande – MS – Brasil

`gabriel.b.gutierrez@ufms.br`

Abstract. *This paper exposes the correlations between weather conditions and the number of fires in Australia. It will explore the impact of temperature and precipitation variations between 2008 and 2017, with the aim of finding correlations between these variables and the number of fires. The study also used machine learning algorithms to try to predict, given a certain weather condition, the number of fires that will occur on a given day.*

Resumo. *Este artigo expõe as correlações entre as condições climáticas e o número de incêndios na Austrália. Nele será explorado o impacto das variações de temperatura e precipitação entre os anos de 2008 e 2017, com o objetivo de encontrar correlações entre essas variáveis e a quantidade de incêndios. O estudo também realizou o uso de algoritmos de aprendizado de máquina para tentar prever, dada certa condição climática, o número de focos de incêndio que ocorrerão em certo dia.*

1. Introdução

Este trabalho tem como objetivo investigar o impacto das mudanças climáticas nos incêndios florestais na Austrália, a fim de responder às seguintes perguntas:

1. É possível prever a quantidade de focos de incêndio em um dia, dadas as suas condições climáticas?
2. Existe correlação entre o aumento da temperatura e a quantidade de focos de incêndio?
3. Como as condições climáticas locais podem afetar a quantidade de incêndios?

Com as respostas para estas perguntas, é possível não só tomar decisões mais bem fundamentadas para a criação de políticas climáticas, mas também organizar respostas mais rápidas aos focos de incêndio e criar planos de prevenção para áreas consideradas de risco.

2. Dados

Para este trabalho foram utilizados dados de diversas fontes, datando de 2008 até junho de 2017. Nesta seção serão tratadas, a fundo, as fontes desses dados e como foram estruturados para utilização durante as pesquisas.

2.1. Datasets

Os dados utilizados neste trabalho são originários de quatro diferentes fontes, quais sejam: MODIS da NASA, Australian Bureau of Meteorology, FAO e Simple Maps.

Os dados referentes aos focos de incêndio e às anomalias térmicas são gerados pelo satélite MODIS da NASA.

Os dados referentes às condições climáticas por cidade foram gerados pelo Escritório de Meteorologia da Austrália (Australian Bureau of Meteorology).

Dados referentes às variações climáticas por ano são dados distribuídos pela FAO e coletados pela NASA-GISS.

Por fim, os dados relativos às coordenadas geográficas de cada cidade são originados do dataset Simple Maps Austrália.

2.1.1 Anomalias Térmicas

O dataset de anomalias térmicas é composto de leituras do satélite modis de 2000 até 2019, com os campos latitude, longitude, brightness, scan, track, acq_date, acq_time, satellite, instrument, confidence, version, bright_t31, frp, day/night e type. Ao todo, o dataset contém 5.081.221 de medições ao longo de 20 anos.

Para este trabalho, contudo, foram utilizados apenas dados de 2008 até 2017, uma vez que esse é o período de intercessão entre todos os datasets utilizados. Também não foram utilizados todos os campos presentes no dataset, mas se considerou a latitude, longitude, brightness, acq_date, confidence e day/night, campos que foram julgados mais importantes por conterem informações únicas e possivelmente úteis aos propósitos deste estudo.

Foi criada uma geofence para delimitar uma área de estudo. Para o trabalho apenas a área entre as latitudes -27.38646 e -43.67288 e as longitudes 153.48215 e 140.48572 foi considerada, uma vez que é nessa área que se concentra a maioria das cidades presentes do dataset de dados meteorológicos. Após a delimitação da área, aproximadamente dois milhões de leituras foram consideradas para o trabalho.

2.1.2. Dados Meteorológicos

O dataset, contendo os dados meteorológicos, tem informações de 2008 até a metade de 2017, com os campos Date, Location, Min Temp, Max Temp, Rainfall, Evaporation, Sunshine, Wind Gust Dir, Wind Gust Speed, WindSor 9am, Wind Dir 3pm, Wind Speed 9am, Wind Speed 3pm, Humidity 9am, Humidity 3pm, Pressure 9am, Pressure 3pm, Cloud 9am, Cloud 3pm, Temp 9am, Temp 3pm, Rain Today, Rain Tomorrow e um total de 145.461 medições de 49 locais diferentes. Para este trabalho foram usados campos date, location, min Temp, max Temp, rainfall, wind Gust Dir, wind Gust Speed e as leituras de 29 cidades.

Os cortes nos campos foram realizados em função da redundância de dados e inconsistência na medição de alguns campos, como evaporation e sunshine. Além disso, Rain Today e Rain Tomorrow foram retirados, uma vez que fogem ao escopo do trabalho.

Também, 20 cidades foram cortadas do estudo, uma vez que se encontravam muito distantes das outras cidades. Após os cortes, foram utilizadas aproximadamente 85 mil medições.

2.1.3. Variação de Temperatura

O dataset contendo a variação de temperatura contém dados sobre mudança de temperatura e desvio padrão mensal, sazonal e anual de 1961 até 2019, para a Austrália como um todo. Como já se mencionou anteriormente, apenas dados de 2008 até 2017 foram utilizados. Os campos Area Code, Months Code, Element Code e Unit foram cortados pela sua insignificância para o projeto. Após as delimitações, obteve-se 350 medições.

2.1.4. Cidades

Como as medições de satélite têm sua localização marcada como coordenadas geográficas, e as medições meteorológicas usam o nome da cidade para ser identificadas foi necessário estabelecer um elo entre os dois dataset. O dataset de cidades fez exatamente isso, ele contém o nome da cidade e as coordenadas geográficas referentes a ela. As cidades utilizadas estão distribuídas da seguinte maneira:



Figura 1. Distribuição das cidades utilizadas no estudo

2.2 Data Warehouse

O data warehouse para este trabalho foi projetado em sql e contém tabelas que refletem os dados dos datasets utilizados.

2.2.1 Modelagem

Como modelagem do data warehouse foi utilizado o seguinte diagrama entidade-relacionamento:

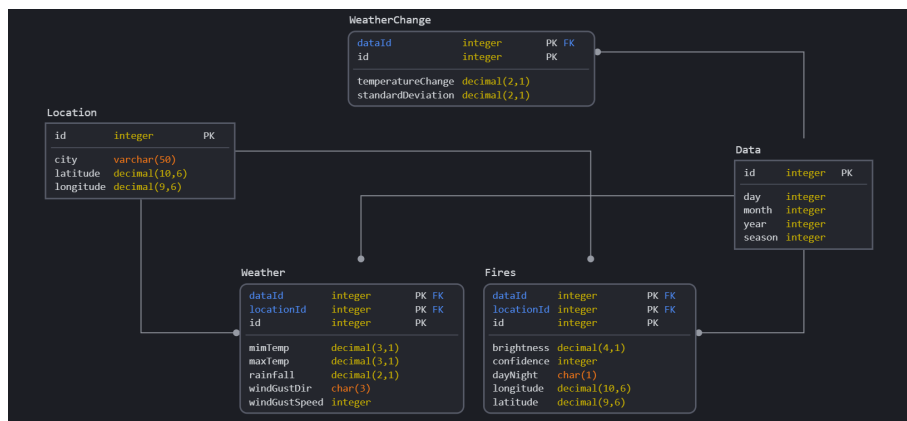


Figura 2. Modelagem do data warehouse

2.2.2 Tabelas Fato

A modelagem inclui as seguintes tabelas fato:

Weather, com as informações de temperatura mínima, temperatura máxima, precipitação, direção do vento e velocidade do vento como fatos.

Fires, com as informações de brilho, confiança, dia/noite, latitude exata, longitude exata como fatos.

Weather Variation, com as informações de mudança de temperatura e desvio padrão como fatos.

2.2.3 Dimensões

O data warehouse conta com as dimensões do tempo, representado por Data, e a do espaço, representado por Localização.

A dimensão Data tem como atributos dia, mês, ano e estação, que é representada por um inteiro seguindo a ordem das estações no hemisfério sul, sendo elas: 1- Verão, 2- Outono, 3-Inverno e 4- Primavera. É importante ressaltar que o verão de um ano começa sempre no mês de dezembro do ano anterior, por isso janeiro de 2017 e dezembro de 2016 são partes do mesmo verão.

A dimensão Localização é representada por cidade e as suas coordenadas geográficas.

Na tabela fato Fires também estão presentes coordenadas geográficas, mas essas são referentes ao foco em si, e não à cidade a qual o foco pertence.

3. Metodologia e Resultados

Como falado anteriormente é preciso fazer a ligação entre a localização dos focos de incêndio e as cidades com as medidas meteorológicas. Para isso, foi usada a fórmula de Haversine, que calcula a distância entre dois pontos no globo, desconsiderando relevo, utilizando a latitude e longitude dos mesmos.

Ao calcular a distância entre cada foco de incêndio e as cidades selecionadas foi alocado a cada incêndio a cidade mais próxima a ele, e assim foi possível correlacionar os dados meteorológicos com focos de incêndios próximos.

É preciso notar, porém, que esse método foi utilizado apenas por não haver como determinar as condições climáticas exatas do local dos focos, o que seria o ideal. Ao utilizar esse método, muitos focos foram indexados a cidades a centenas de quilômetros de distância de onde realmente aconteceram, uma vez que não havia cidades mais próximas. Isso pode ter causado ruído e imprecisão nos resultados que serão relatados a seguir.

Para a apresentação dos resultados serão utilizadas como referência quatro cidades: Moree, Brisbane, Wagga Wagga e Cobar. Essas cidades foram escolhidas uma vez que são as quatro cidades com o maior número de focos de incêndios indexados a elas.

3.1 É possível prever a quantidade de focos de incêndio em um dia, dadas as suas condições climáticas?

Prever a quantidade de incêndios é uma tarefa com muitas variáveis. A quantidade de dados disponíveis, as variáveis usadas na previsão e o algoritmo utilizado são algumas delas. Estas foram os pontos tratados com maior cuidado.

Para chegar a um resultado consideravelmente satisfatório, foram necessárias várias iterações tanto de informações dadas aos algoritmos, quanto a busca do melhor algoritmo em si. Na busca pelos melhores algoritmos, foram utilizados os dados meteorológicos de todas as cidades. Foram, também, testados, ao todo, sete algoritmos, sendo eles: MLP Regressor, Linear Regression, SGD Regressor, Elastic Net, Bayesian Ridge, Gradient Boosting Regressor e SVM. Ao final dos testes, os dois algoritmos com melhor desempenho foram selecionados para a fase de testes subsequentes.

Tendo encontrado os melhores algoritmos para a tarefa, MLP Regressor e Linear Regression, era necessário achar os melhores parâmetros para esses algoritmos. Para isso foi utilizado o Grid Search para testar múltiplas combinações de parâmetros e comparar os resultados entre elas. Para ambos os algoritmos foram utilizadas as variáveis temperatura mínima, temperatura máxima, precipitação, direção dos ventos, velocidade dos ventos, cidade e um histórico de 3 dias contendo a quantidade de focos registrados nesta janela de tempo.

Para o MLP Regressão foram testados os tamanhos das camadas, a função de ativação e o número máximo de iterações do algoritmo. Foram utilizados os seguintes valores:

- hidden_layer_sizes: (50,100,200,400,200,100,50,), (50, 100, 50,), (25,11,7,5,3,), (50, 100, 150, 150, 100, 50,)
- activation: relu, tan, logistic, identity
- max_iter: 500, 1000, 1500

Para o Linear Regression foram testados os parâmetros: fit intercept, normalize, copy X e positive com os valores variando entre verdadeiro ou falso.

Essa etapa foi a mais difícil computacionalmente falando, uma vez que foi preciso treinar todas as combinações de parâmetros com uma base de dados de mais de 77 mil medições. Após aproximadamente 40 horas de processamento, pôde-se determinar a melhor combinação de parâmetros para cada um dos algoritmos.

A melhor combinação de parâmetros para o MLP Regressor foram hidden layer sizes: (25,11,7,5,3,), activation: logistic com um máximo de ativações de 100, tendo um erro quadrático médio de 0.7, e um R^2 de 0.2.

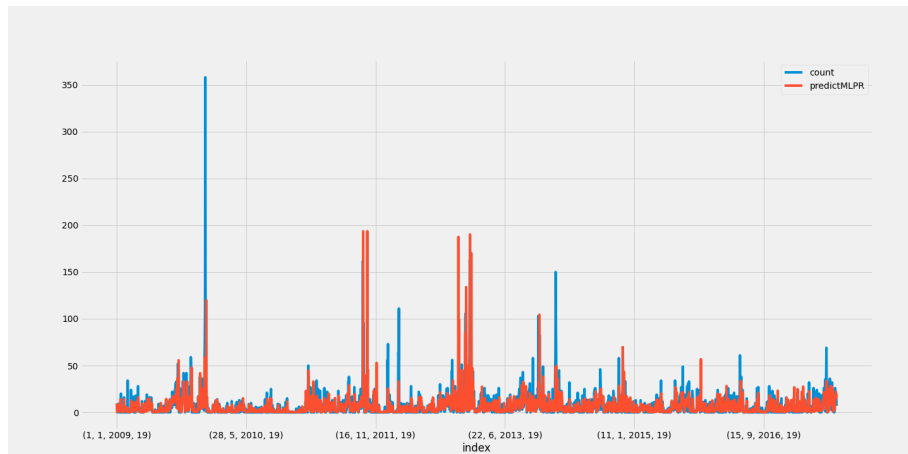


Gráfico 1. Contagem real x Previsão Moree (MLP Regressor)

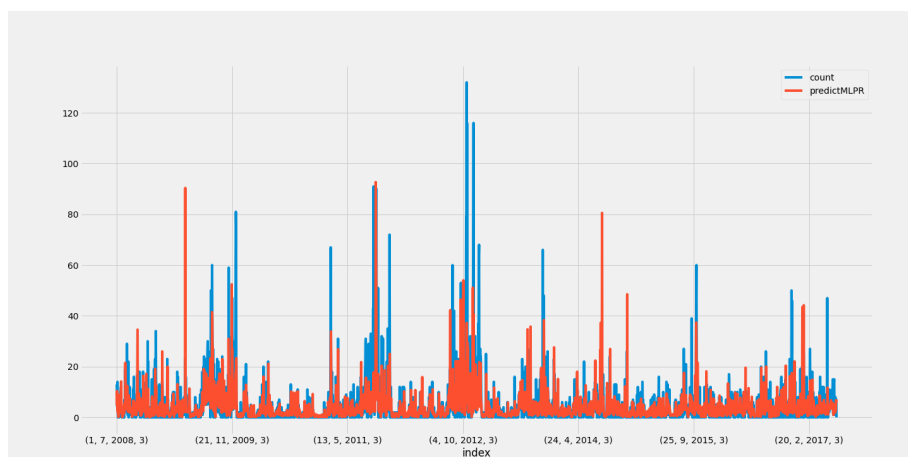


Gráfico 2. Contagem real x Previsão Wagga Wagga (MLP Regressor)

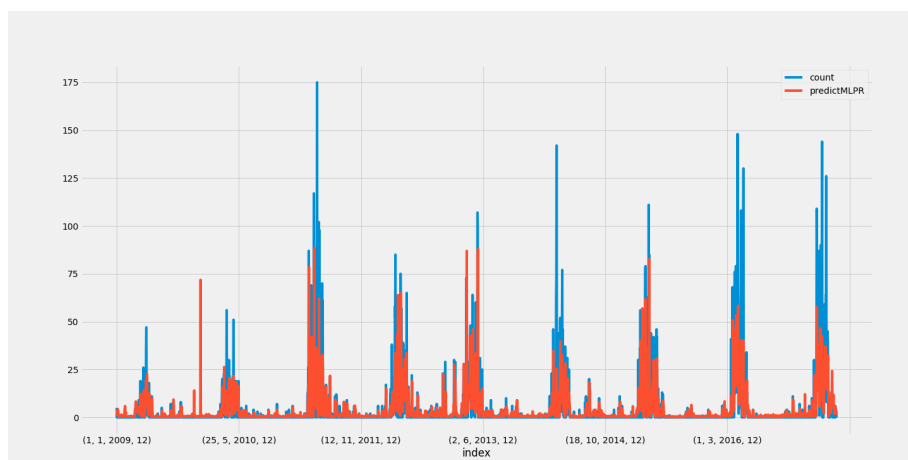


Gráfico 3. Contagem real x Previsão Brisbane (MLP Regressor)

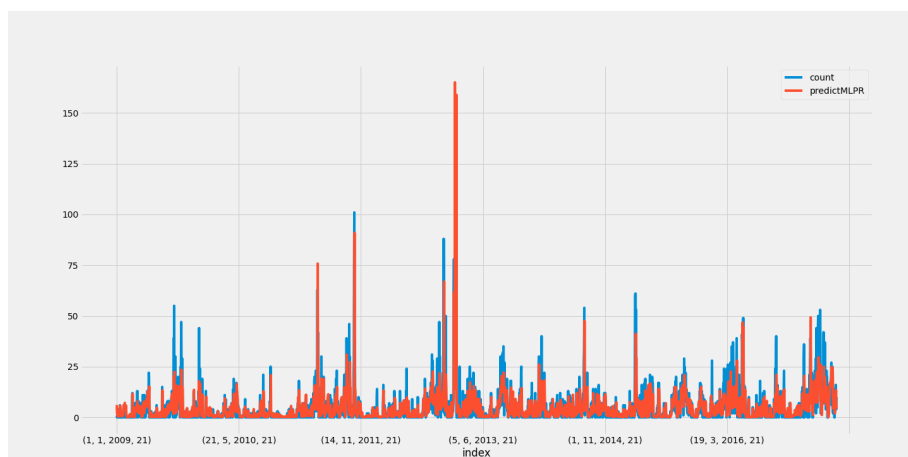


Gráfico 4. Contagem real x Previsão Cobar (MLP Regressor)

Para o Linear Regression, a melhor combinação foi copy X e normalize ativados e fit intercept e positive desativados, com um erro médio quadrático de 0.69 e um R^2 de 0.30.

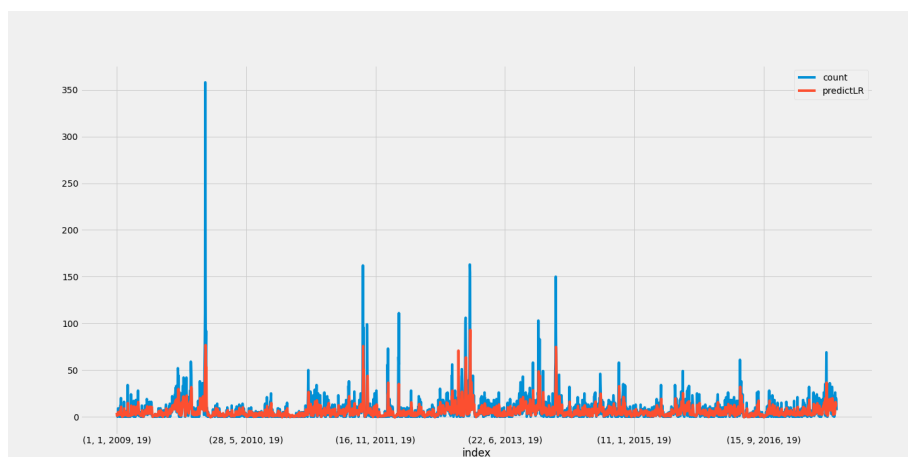


Gráfico 5. Contagem real x Previsão Moree (Linear Regression)

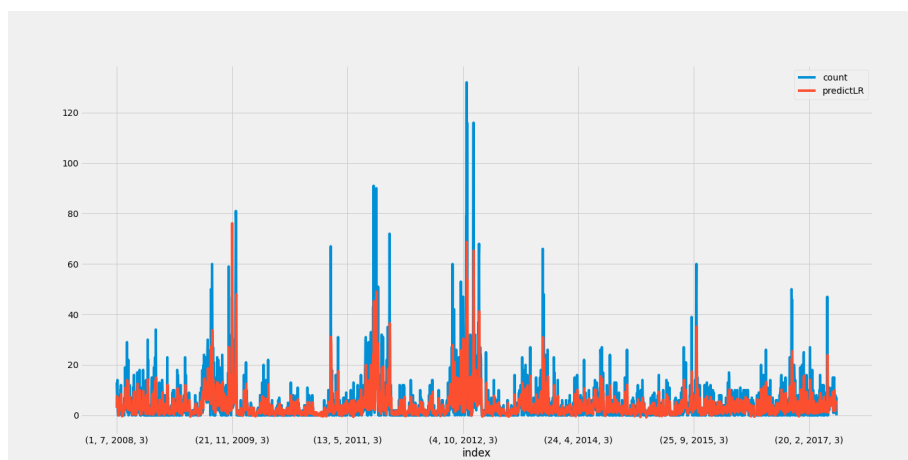


Gráfico 6. Contagem real x Previsão Wagga Wagga (Linear Regression)

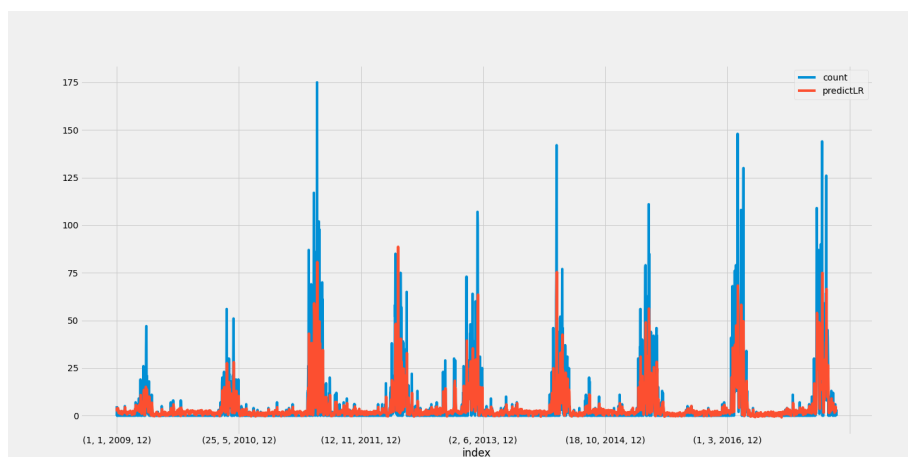


Gráfico 7. Contagem real x Previsão Brisbane (Linear Regression)

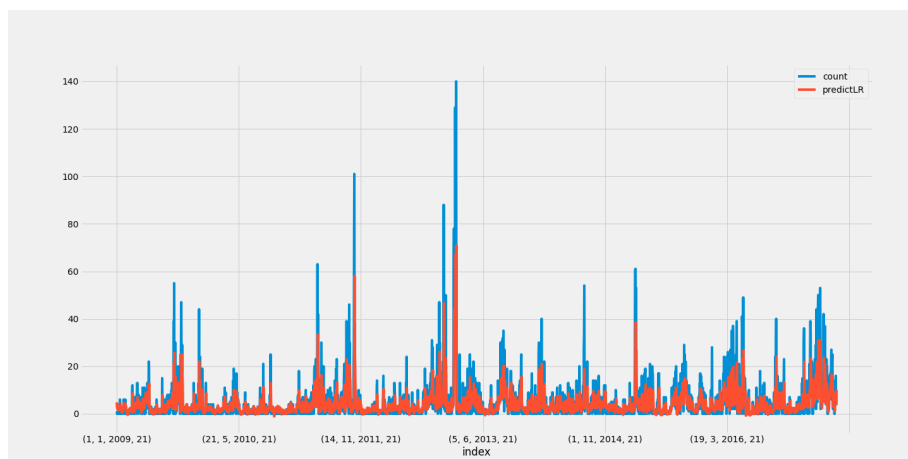


Gráfico 8. Contagem real x Previsão Cobar (Linear Regression)

Mesmo faltando precisão ao prever quantos incêndios exatamente acontecerão dado as condições climáticas do local é possível observar que ambos conseguem prever variações na quantidade de incêndios em um dia. Essa habilidade, mesmo que imprecisa no quesito de quantidade, é valiosa na hora de tomar decisões de planejamento e resposta a combate de incêndios, uma vez que oferece a chance de responder mais adequadamente à determinada situação.

É possível que uma melhora na previsão dos algoritmos seja alcançada com a utilização de dados meteorológicos mais precisos, uma vez que, como foi dito anteriormente, vários focos foram alocados a cidades muito distantes de onde eles realmente aconteceram. Isto significa que as condições climáticas usadas para representar a sua ocorrência não sejam, necessariamente, o que de fato ocorreu.

3.2 Como as condições climáticas locais podem afetar a quantidade de incêndios?

Para responder a esta pergunta foram utilizados os dados dataset e compilados nos tópicos apresentados a seguir.

3.2.1 Temperatura média

Os gráficos que seguem representam a média móvel da temperatura média de cada uma das cinco cidades e o número de incêndios em um determinado dia. Para calcular a média móvel foi utilizada uma janela de 15 dias, buscando uma suavização da variação da temperatura e do número de incêndios de cada dia e ,assim, facilitando a visualização dos dados.

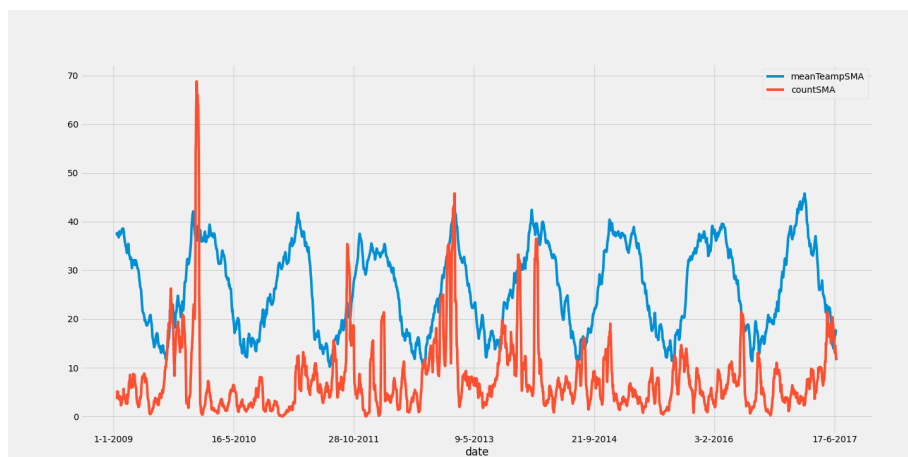


Gráfico 9. Temperatura x Incêndios Moree

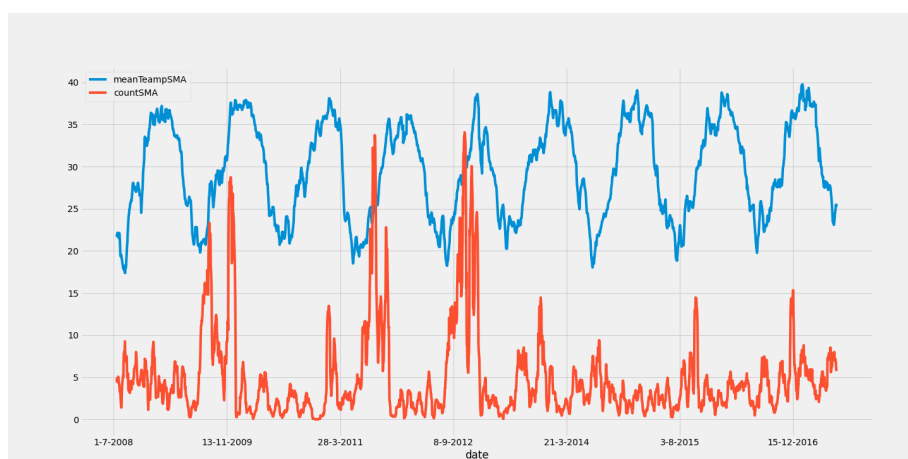


Gráfico 10. Temperatura x Incêndios Brisbane

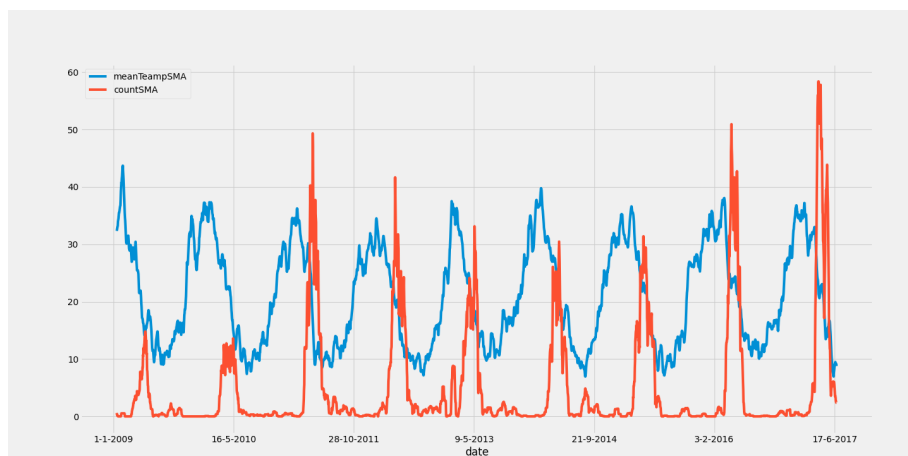


Gráfico 11. Temperatura x Incêndios Wagga Wagga

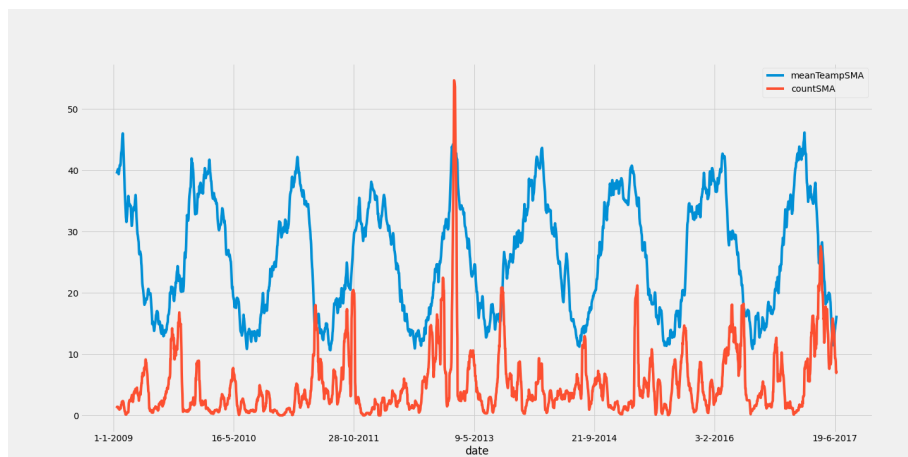


Gráfico 12. Temperatura x Incêndios Cobar

Em todos os gráficos é possível ver o aumento da temperatura com a chegada do verão. Nesta estação, o número de focos de incêndio sobe drasticamente, sugerindo alta correlação entre temperatura e o número de focos.

3.2.2 Precipitação

Diferentemente da temperatura, a precipitação não segue um padrão tão definido durante os anos, porém é visível que as maiores concentrações de focos acontecem quando o nível de precipitação é muito baixo ou até mesmo zero.

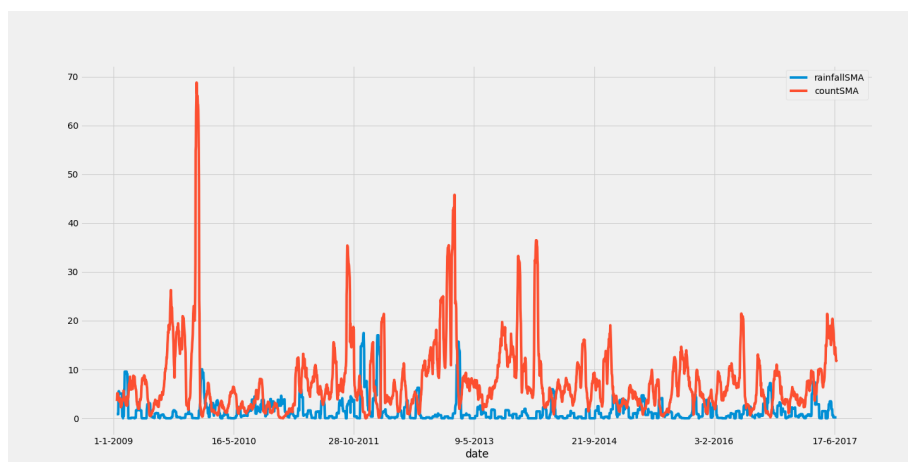
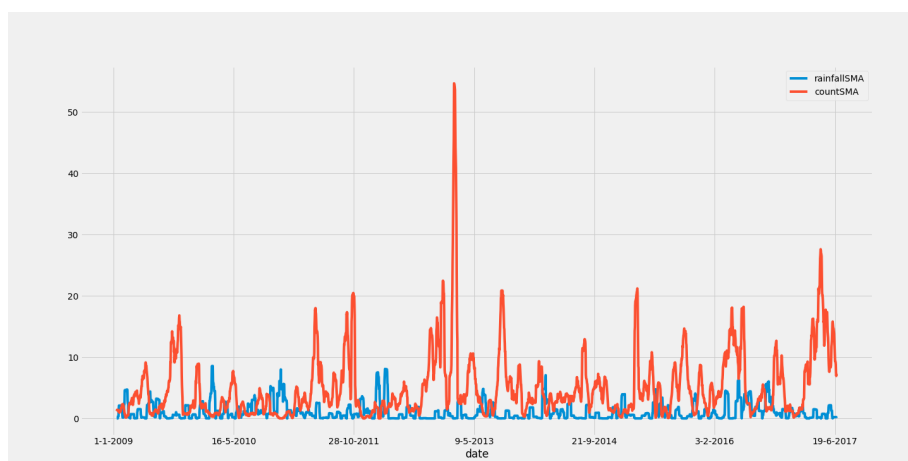
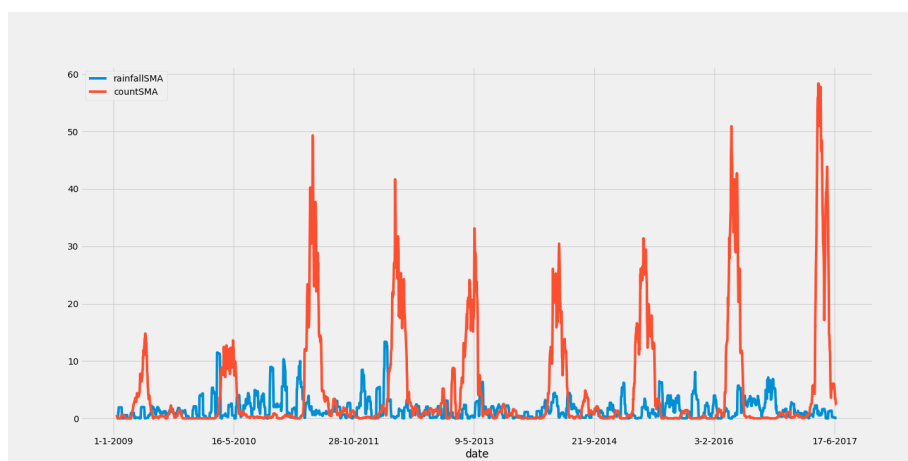
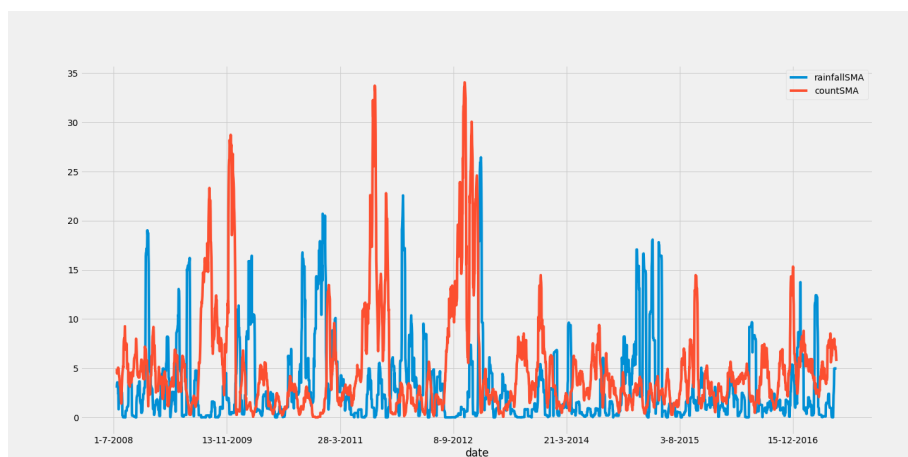


Gráfico 13. Precipitação x Incêndios Moree



3.3 Existe correlação entre o aumento da temperatura e a quantidade de focos de incêndio?

Para entender melhor se há relação entre a variação de temperatura e a quantidade de incêndios, essas duas informações foram plotadas em dois gráficos, um com informações anuais e outro com dados mensais.

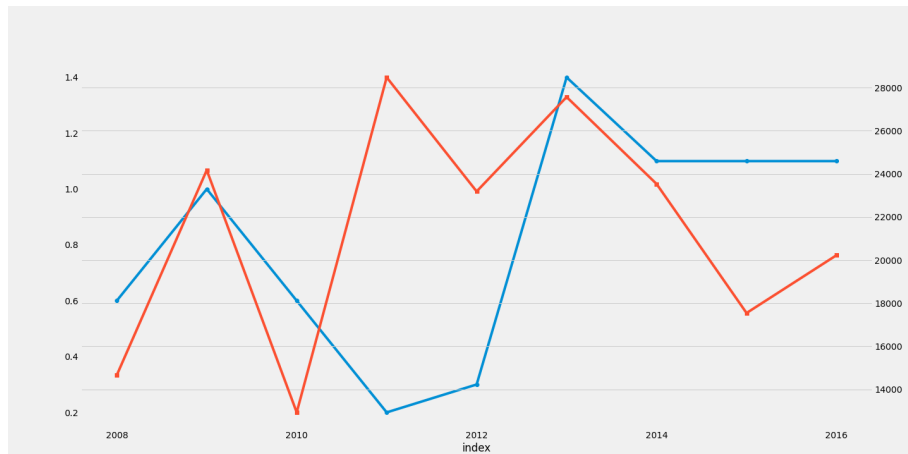


Gráfico 17. Variação de temperatura x Incêndios anual.

No gráfico 17, temos plotado em azul a variação de temperatura e em laranja a quantidade de incêndios. Em algumas instâncias é possível ver a quantidade de incêndios acompanhado o aumento ou diminuição da temperatura, porém também é possível observar o contrário acontecendo.

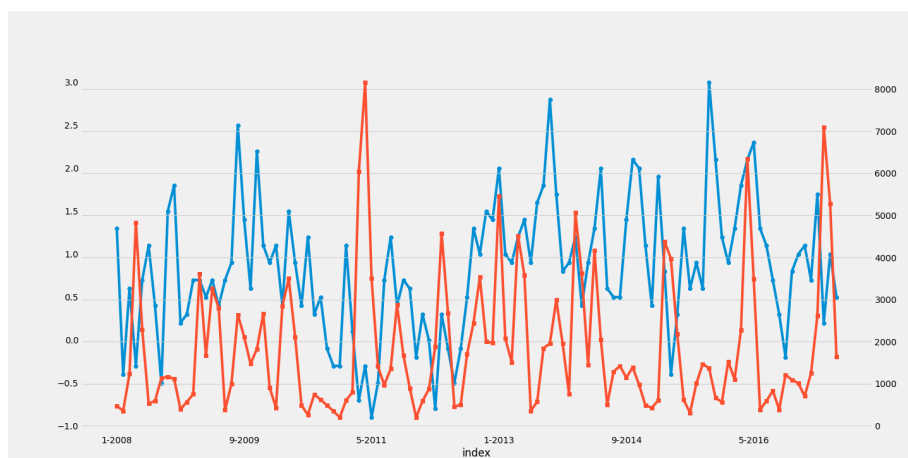


Gráfico 18. Variação de temperatura x Incêndios mensal.

Considerando uma escala mensal, o mesmo comportamento foi encontrado, porém é possível perceber mais facilmente a correlação entre a variação da temperatura e a quantidade de focos de incêndio. Em quase todas as instâncias, quando a temperatura varia, seja para mais ou para menos, a quantidade de incêndios acompanha. Na maioria das vezes essa variação não tem a mesma intensidade, entretanto ela está presente. Assim sendo, podemos concluir que a temperatura tem impacto na quantidade de focos de incêndio, principalmente se levado em conta também os dados mostrados na seção 3.2.

4. Conclusão

Este estudo mostrou a correlação entre as condições climáticas e a quantidade de focos de incêndio na Austrália. Os dados evidenciam que as altas temperaturas ensejam maior quantidade de incêndios. Essa observação merece atenção, visto que o aumento da temperatura global nos próximos anos é uma tendência, e o aumento do número de incêndios provocados pelas altas temperaturas coloca em risco os ecossistemas locais.

Paralelamente, foi desenvolvido um método para prever a quantidade de incêndios em um determinado local, dadas as suas condições climáticas. Esse método, mesmo carecendo de precisão, é definitivamente efetivo em prever flutuações na quantidade de incêndios de um local, possibilitando, assim, as tomadas de decisões mais bem informadas por parte das autoridades locais, o que resultaria em maior eficiência no combate aos focos de incêndio.